Count regression practice: Asthma-related Emergency Room Visits

STAT 245

Data

The data you will use today are stored at: https://sldr.netlify.app/data/asthma.csv. All data are from the metro Detroit area, and include demographic and location information along with information about asthma-related hospital emergency room visits.

The data are from several sources and there is slight temporal mismatch among variables: - Variables `ERVisits_Asthma_0_17`, `median_income`, and `TotalMedicaidPop0_17` are from data.gov. The ER visit data are from 2013. The income and Medicaid data are averaged over the period 2010-2014, based on data from the US Census' American Community Survey. - All other variables are from 2010 US Census data, which I accessed via the R package `zipcodeR`.

Variables in the dataset are:

- `Zip`: ZIP code (postal code)
- `ERVisits_Asthma_0_17`: Number of asthma-related hospital emergency room visits by kids between 0-17 years old, for the year 2013
- `TotalMedicaidPop0_17`: Number of children 0-17 with medical insurance provided by Medicaid
- `median_income`: Median household income, in $10,000s
- `major_city`: Name of the city the zip code is located in
- `county`: Name of the county the zip code is located in
- `state`: State (always Michigan)
- `latitude` in decimal degrees North
- `longitude` in decimal degrees East
- `population` of the zip code (number of individual residents)
- `land_area_in_sqmi`: area within the zip code (in square miles) that is not covered by water
- `water_area_in_sqmi`: area within the zip code (in square miles) that is covered by water
- `occupied_housing_units`: number of occupied buildings in the zip code
- `median_home_value`: median value of homes in the zip code, in $10,000s

Tasks

1. Plan a Model

Make a plan for a regression model with `ERVisits_Asthma_0_17` as the response variable.

You can choose your own research question (and key predictor variable to consider).

Your goal is to try to understand factors related to the number of ER visits for asthma in this area of Michigan. This exploration will contextualized a bit (soon!) by an upcoming Moodle Forum activity on industrial air pollution hot-spots. But for now…you get some practice using count regression models.

As you plan…Be sure to consider the size of the dataset, and include a causal diagram in your process. Consider: are there any interactions?

2. Fit your Model

Skip data exploration and graphics in the interest of time - return to it at the end if you have extra time, though.

Use `glmmTMB()` (don't forget to run `library(glmmTMB)` first) to fit the count regression model you have planned, and view the model `summary()`.

You are not required to write down the model equation (although you can for practice, if time permits).

Check with your prof in case of any worrisome warning or error messages!

3. Do model assessment

Use graphical model assessment to check the conditions (listed below). For each one, make the graph(s) you need, state whether you think the condition is met, and provide a rationale based on what you see in the graph.

- (log)-linearity of predictor-response relationships
- independence of residuals
- mean-variance relationship

Remember, to check the mean-variance relationship, you will use the functions `simulateResiduals()` and `plotResiduals()` from the package `DHARMa` to re-scale the residuals and graph the scaled residuals vs fitted values.

4. Extra Time?

Options to fill your surplus time:

- Do any necessary project work
- Make graphs of the data (we skipped this step earlier in the interest of time)
- Try to do ***model selection*** (using ANOVA or information criteria) or ***make prediction plot(s)*** for your model – for both, the process does not change for count vs. linear models, but more practice is always good!) **What conclusions can you draw, once you have the assessment results, selection results, and prediction plots?**